Project Design Phase -I

Proposed Solution Template

Date	25 September 2022
Project Name	Flight Delay Prediction

Proposed Solution Template

S. No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	As discussed, weather condition plays an important role in proper and timely functioning of flights. We propose a flight delay prediction system which focuses mainly on predicting delay of a flight based on the weather situation. To make the system more scalable it is necessary to choose an algorithm which considers all the parameters to be independent. Supervised learning as the name indicates a presence of supervisor as teacher.
2.	Predictive Modelling on Weather Data	• The dataset contains numerous factors that we might want to bar. The Taxi in/out variables and all the delay variables are to be dropped. We are most interested in the weather variables, so we decide to select only them from the dataset. • The new dataset contains variables for the scheduled operation at the airport - more or less an indicator of the demand per hour in the airport. This indicator is broken down into departures operations (flights per hour) and arrivals operations - both

		would count in determining the demand in any given hour at the airport, so we combine them into a sum of airport 'demand'.
3.	Prediction model	• Airline passengers to know whether the flight will get delayed or not. To make the system more scalable it is necessary to choose an algorithm which considers all the parameters to be independent. • The name indicates a presence of supervisor as teacher. Essentially supervised learning could be a learning that within which we tend to teach or train the machine exploitation data which is well tagged which means some data is already labelled with correct answer.
4.	Data Pre-processing	It is required to carried out the pre-processing steps on the flight dataset, before implementing the machine learning models. The following shows some pre-processing techniques. • **Feature Encoding:** In this research, feature present in the Airline dataset such as Airline, Airport From, Airport To are categorical values and it cannot be processed directly. Before applying traditional machine learning models, it must be converted into numeric values. • **Data Normalization:** Feature with different scale can heavily affect the performance of the machine learning model. Standardization technique is utilized to perfume scaling on the features values of given dataset.

5.	Gradient Boosting Ensemble Methods	• Based on the performance and popularity among various machine learning algorithms, gradient boosting trees (GBTs) based algorithms are well known for the structured data. In this research following are the three GBT based algorithms are selected to predict the flight delay.